

AD-A152 950

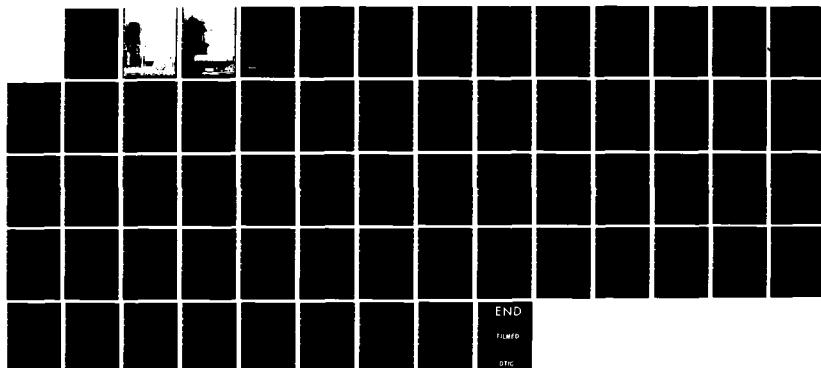
CLARIFY (TRADEMARK): AN ON-LINE GUIDE FOR REVISING
TECHNICAL PROSE(U) RAND CORP SANTA MONICA CA
M E VAIANA ET AL. NOV 83 RAND/N-2037-RC

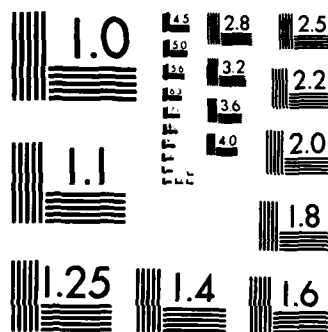
1/1

UNCLASSIFIED

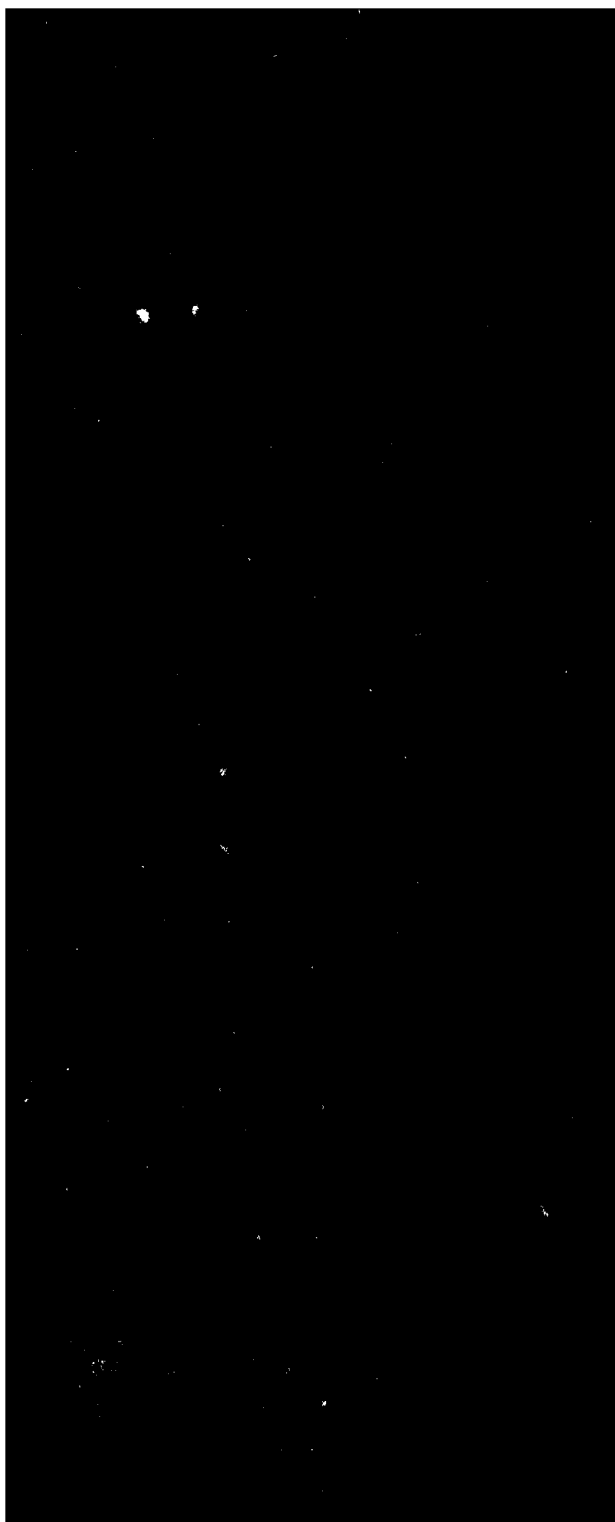
F/G 5/7

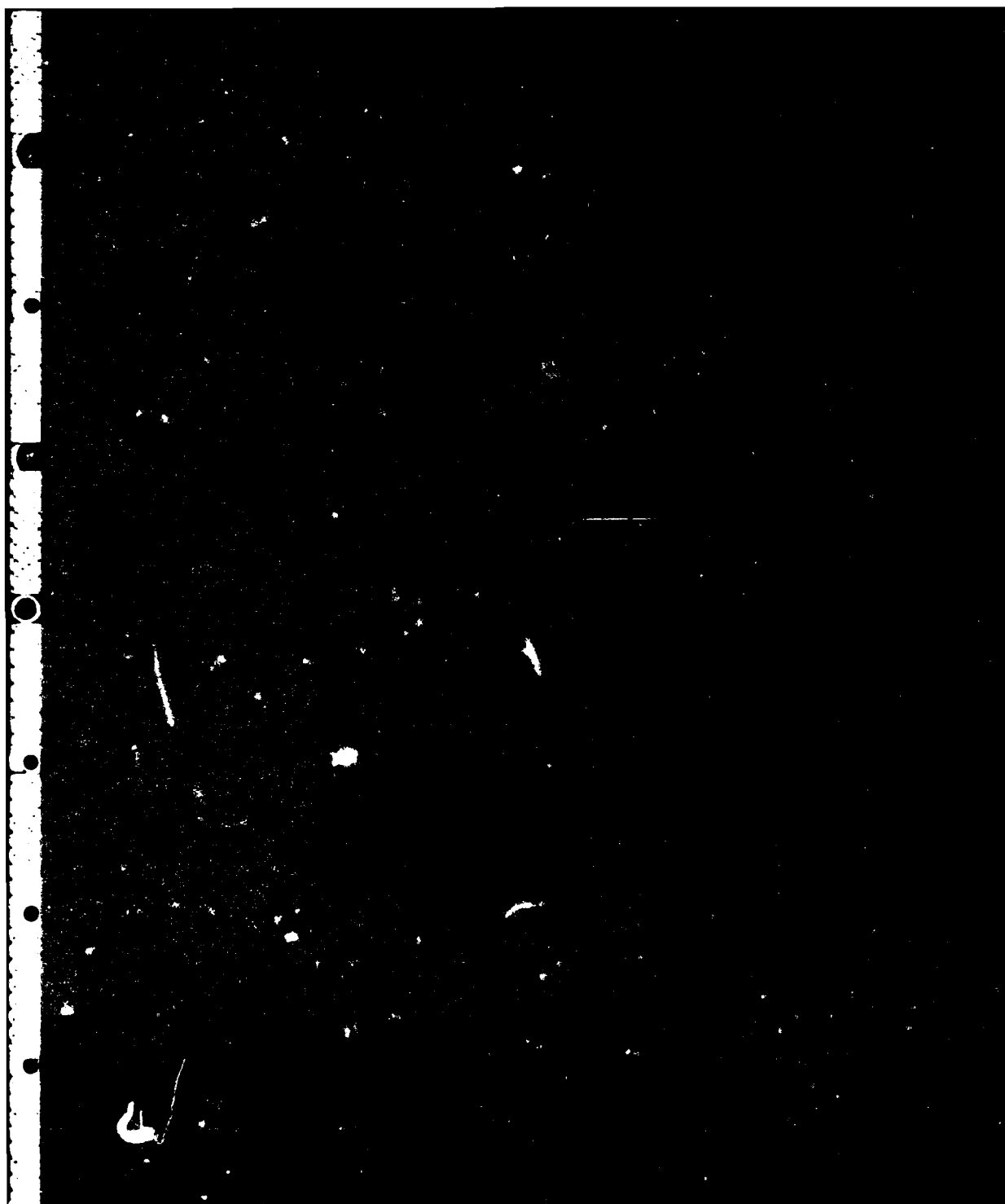
NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A





A RAND NOTE

CLARIFY:TM AN ON-LINE GUIDE FOR REVISING
TECHNICAL PROSE

Mary E. Vaiana, Norman Shapiro,
Mark LaCasse

November 1983

N-2037-RC

A CLARIFYTM Report

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-2	



35th
Year



PREFACE

This Note describes the development and initial testing of CLARIFY™,¹ an on-line writing aid designed to guide the revision of technical prose. Funding for this research was provided by The Rand Corporation. The Note should be of interest to linguists, psychologists, document designers, teachers of writing, and others concerned with using computer technology to improve communication. A subsequent report will present the results of the larger-scale testing of CLARIFY that is now in progress.

¹CLARIFY is the trademark and service mark of The Rand Corporation.

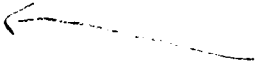
SUMMARY

This Note describes the development and testing of CLARIFY, a computerized writing aid designed at The Rand Corporation to assist writers in revising technical prose. CLARIFY is not a traditional readability formula; its design reflects research on how English speakers go about the task of understanding sentences.

CLARIFY flags sentences that have certain patterns of nominalizations, prepositional phrases, and forms of the verb *to be*. The choice of these features reflects research which suggests that the dominant strategy employed by English speakers in interpreting sentences is the assumption of a subject-verb-object (SVO) structure. The features that CLARIFY flags are good surrogate indicators that a sentence does not have an SVO structure, and therefore that the initial interpretive strategy will be unsuccessful. In developing CLARIFY, we tested various patterns of these features and obtained user comments about the system's usefulness and effectiveness.

Like all computerized writing aids, CLARIFY has limitations. The most important are (1) it functions only at the sentence level, and (2) it uses surrogate rather than directly causal measures of comprehension. Despite these limitations, the test users found that CLARIFY prompted them to revise more extensively, more quickly, and more effectively.

Authors can work with CLARIFY output either on-line or in hard copy. CLARIFY is in general use at The Rand Corporation, where it is also continuing to be tested.



ACKNOWLEDGMENTS

We are indebted to our Rand colleagues Carl Builder, Edward Merrow, and Joyce Peterson for comments on earlier drafts of this Note. Janice Redish, Director of the Document Design Center, American Institutes for Research, provided an insightful technical review, and Janet DeLand a crisp and skillful editing. Among the initial test users, Dennis De Tray was particularly generous with time and provocative suggestions. We are also grateful to staff at the Document Design Center for their extremely helpful comments early in the system's development, and to Mary Anderson and Connie Greaser, of The Rand Corporation, for supporting our efforts in a variety of ways.

CONTENTS

PREFACE	iii
SUMMARY	v
ACKNOWLEDGMENTS	vii
Section	
I. INTRODUCTION	1
II. A BRIEF HISTORY OF READABILITY FORMULAS	3
Early Readability Formulas	3
Computerized Readability Formulas	4
Why Readability Formulas Fail	8
III. THE CLARIFY SYSTEM: RATIONALE AND DESIGN SPECIFICATIONS	12
Designing an On-Line Revision Guide	12
The Role of the Sentence in Understanding Text	14
Selecting Surrogate Features for CLARIFY	20
IV. USING THE CLARIFY SYSTEM	25
How CLARIFY Works	25
Initial Tests of CLARIFY	31
Evaluating Effectiveness	34
Assessing CLARIFY	37
V. CURRENT TESTING PLANS	40
Research Questions	40
User Groups	41
Evaluation Procedures	43
Appendix	
A. Suggested Guidelines for Evaluating CLARIFY	45
B. Questionnaire Used to Evaluate CLARIFY in Interview with Test Users	46
C. Questionnaire on Composing Styles	48
REFERENCES	51

I. INTRODUCTION

This Note describes CLARIFY, a computerized writing aid developed at The Rand Corporation to assist writers in revising technical prose. The system differs from traditional "readability formulas" both in basic concept and in implementation. CLARIFY is based on extensive research in linguistics and cognitive psychology on how humans understand and store information gained from individual sentences.

CLARIFY flags sentences that have certain patterns of nominalizations,¹ prepositional phrases, and forms of the verb *to be*. Our choice of these features was based on research which suggests that the dominant strategy employed by English speakers in interpreting sentences is the assumption of a subject-verb-object (SVO) structure. Sentences that have this structure are found to be most easily understood. This research also indicates that nominalizations, prepositional phrases, and forms of the verb *to be* are good surrogate indicators that a sentence does not have an SVO structure, and therefore that the initial interpretive strategy will be unsuccessful. In developing the CLARIFY system, we tested various patterns of these features, using a large database of sentences written by Rand staff members. We also asked a group of users to provide comments on the usefulness and effectiveness of the system.

Like all on-line writing aids, CLARIFY has its limitations. For example, it does not flag some sentences that are difficult to understand and should be revised. Moreover, although CLARIFY provides cues and principles to assist in revising sentences, successful use of the system ultimately depends on the author's skill at that task. In its current form, CLARIFY functions only at the sentence level. And because CLARIFY does not parse sentences, it must use surrogate features to represent the causal measures of sentence comprehension.

¹Nouns formed by adding one of several suffixes (such as -ment, -ence, -tion) to the stem of a verb or adjective, e.g., *confinement*, *intelligence*, *aggravation*.

Despite these limitations, all of our test users stated that CLARIFY prompted them to revise more extensively, more quickly, and more effectively. They also found that regular use of the system resulted in very beneficial teaching effects.

The background, development, and prototype testing of the CLARIFY system are discussed in the following sections. Section II summarizes the history of readability formulas and discusses applications in which they have been used inappropriately. Section III describes the development of the CLARIFY system and the research on which it is based. We examine the role of the individual sentence in text comprehension and show how CLARIFY reflects the dominant sentence-mapping strategies of English speakers. Section IV describes how CLARIFY works and presents a summary of comments from test users concerning the system's effectiveness. Section V discusses current plans for testing and evaluating CLARIFY. Finally, the questionnaires used in our initial evaluations of CLARIFY are reproduced in the appendixes.

II. A BRIEF HISTORY OF READABILITY FORMULAS

EARLY READABILITY FORMULAS

Traditional readability formulas, especially computerized versions of such formulas, are often used to guide revision of prose text. However, they were not designed for this purpose, and they should not be used for it.

The context in which readability formulas evolved and the applications for which they were originally designed have been summarized by Harris and Jacobson (1979), Klare (1975), and Redish (1980).

A readability formula counts certain language variables in a piece of text to provide an index of the text's reading difficulty level. The index of difficulty is based solely on certain text features; it does not involve any assessment of the actual difficulty that readers experience with the text. Readability formulas are therefore primarily useful for determining whether textbooks and training manuals are too difficult for their intended audiences.

The first easy-to-apply readability formula was published by Irving Lorge in 1939. Designed for grades 3 through 12, its variables were the number of words in a sentence, the number of prepositional phrases per 100 words, and the number of difficult words that did not appear on a specified list (a list of 3000 words devised by Dale and Chall, 1948). In later versions of the Lorge formula, these variables were simplified.

Perhaps the best known readability formula was designed by Rudolf Flesch in 1943 to assess the difficulty of general adult reading material. The formula, which Flesch called the Readability Ease Formula, has appeared in many forms. The most widely used version considers the number of syllables per 100 words and the average number of words per sentence. The formula has been used extensively and has been converted essentially unaltered into computerized form (Coke and Rothkopf, 1970).

Another widely used formula, developed by Dale and Chall in 1948, has as its variables the average sentence length in words and the percentage of words outside the specified list noted above. Klare (1963) suggests that this was the best general-purpose formula in existence up to 1960.

These early researchers set the precedent for subsequent work on readability formulas. Most researchers have continued to emphasize sentence length and measures of vocabulary as the variables of interest and have focused their efforts on speed and ease of calculation. There are, of course, exceptions. Jacobson (1965) developed formulas for specific kinds of texts (high-school and college physics and chemistry texts), and several researchers have proposed measures of syntactic complexity (Bormuth, 1969; Aquino, n.d.; Botel and Granowsky, 1972a,b; Coleman, 1968; and Selden, 1977).

COMPUTERIZED READABILITY FORMULAS

The computer makes readability formulas easier and faster to use, and since 1960, researchers have been creating automated versions of their own formulas and those of others (see Klare, 1974). These computerized versions are basically straightforward translations of the traditional variables, with neither the formulas nor the approach rethought.

In a few recent, sophisticated applications, readability formulas have been integrated into other assessments of text. Two of these new applications, the Writer's Workbench and the EPISTLE program, are described briefly below.

The Writer's Workbench was developed over a period of years at Bell Laboratories (Murray Hill and Piscataway, N.J.).¹ It consists of a set of 32 computer programs that perform proofreading and stylistic analysis, and provide on-line reference information on English usage and Writer's Workbench programs.

¹See *Discover*, July 1981; *Editor and Publisher*, April 4, 1981; *Business Technology*, July 1983; Coke, 1982; Macdonald, 1982.)

The Writer's Workbench assesses readability by the use of a program called STYLE (Cherry, 1981, 1982; Macdonald, 1982). STYLE provides information about the average lengths of words and sentences, the distribution of sentence lengths, the grammatical types of sentences used, the percentage of verbs that are in the passive voice, the percentage of nouns that are nominalizations, and the number of sentences that begin with expletives. STYLE also calculates four readability formulas: the Kincaid Formula, the Automated Readability Index, the Coleman-Liau Formula, and a version of the Flesch Reading Ease Formula. All of these are traditional formulas that use measures of sentence and word length to determine readability, and they characterize the difficulty of the text in terms of these measures.

STYLE output is provided in both tabular and interpretive form. The tabular format is designed for research purposes and is quite difficult to interpret (Macdonald, 1982). A sample is shown in Fig. 1 (Cherry, 1981, p.2).

The STYLE program does not interpret the statistics it generates, nor does it suggest specific changes. In the STYLE and DICTION programs, "sentence type," "word usage," and "sentence opener" measures are designed to call attention to "overuse of particular constructions" (Cherry, 1981). The program documentation suggests that the user may want to transform some of the overused constructions "into another form" (Cherry, p.5) to vary the sentence structure and length and thereby avoid monotony. This advice reflects the guidelines that writing experts (e.g., Strunk and White, 1959) have provided for writers of *literary* prose. However, these guidelines may not be appropriate for writers of *technical* prose (Coke, 1982). Short, simple sentences may convey technical information more effectively than a varied mix of long and short sentences. And there are no empirical studies to suggest that varied sentence types facilitate or improve comprehension.

The PROSE program provides both statistics and an interpretation of them. PROSE compares the features identified by STYLE to a set of standards and gives a general characterization--in English--of the text. An example is shown in Fig. 2.

To investigate this hypothesis, the researchers presented native English speakers with unnatural situations in which the forms and functions that usually occur together compete. They asked them to interpret sentences that varied word order, animacy, topicalization, and contrastive stress. Afterwards, they asked the subjects what factors influenced their decisions about the sentences. The sentences were structured so that sometimes all the sources of information converged on the same interpretation--for example, word order and animacy might both signal a surface subject. Other test sentences pitted sources of information in competition, such as word order vs. animacy.

Their basic findings were the following:

1. Word order and animacy are the major factors determining sentence interpretation. Topicalization and stress are weaker factors that usually ally themselves with word order or animacy.
2. Tests in which word order and animacy compete show that English speakers rely much more heavily on word-order information than on the semantic information provided by the individual words.
3. Responses were clearer and faster when information from all sources converged on a single interpretation.
4. Respondents seemed to be aware of their respective hierarchy of mapping strategies.

The dominant SVO mapping strategy produced fast reaction times and consistent responses "even in the face of conflicting information from lexical items" (p. 294). The best possible convergence for English speakers is for word order to converge with animacy: noun (animate)-verb-noun (inanimate). Sentences that match these prototypes are easily processed. To the degree that lexical and syntactic information conflict, processing resources that might have been allocated to other aspects of comprehension--at both the sentence and the text level--must be spent on sorting out an interpretation.

These findings from cross-linguistic analysis are consistent with earlier studies, especially those that contrast memory and comprehension

"topic" and "agent."⁵ New instances--that is, elements of new sentences to be interpreted--are assigned a surface grammatical category on the basis of their resemblance to the prototype, ranging from "best" instances to those in which the matching becomes quite fuzzy. In English, the element that provides the best fit to the topic-agent category is usually assigned the surface role of subject. For example:

comment topic-agent

↓ ↓

Amanda loved her calico cat. The cat kept the birds away
from her garden.

When high probability overlaps, such as topic-agent breakdown, languages must provide alternative ways to express the same functions separately. Thus, even though the best fit for the surface subject of declarative sentences in English is the topic-agent, sometimes we want to topicalize the object of the verb and also identify the agent. In English we do this by assigning the surface subject role to the topic-object and identifying the agent with a "by" phrase. Verb agreement goes with the subject, as always. For example:

comment topic

↓ ↓

Amanda loved her calico cat. The cat was always examined
by the veterinarian at the slightest indication of sickness.

↑

agent

This model predicts that sentences with grammatical subjects close to the prototype should be analyzed faster than sentences with less prototypical subjects, or sentences that have a number of competing interpretations.

grammatical subject, while in the second sentence, the subject is "expert."

⁵The *topic* is the "known" information in a sentence that usually appears at the beginning of the sentence and provides context. In contrast, the *comment* is the new information that usually appears at the end of the sentence. The *agent* is the animate noun that performs the action expressed by the verb.

In sentence (C), the semantics are plausible--performers are often sent flowers--but an SVO interpretation is not appropriate. In (D), the interpretation is less plausible--the performer is sending the flowers--but an SVO interpretation is correct.

This study also provides strong evidence of the centrality of the verb in sentence processing. The empirical results suggest that the reader assigns the structurally preferred analysis of a sentence (SVO), then uses real-world knowledge to consider the possible sets of relations between phrases to see if the initial structural analysis can be supported. The most important clue to those relations is the verb, because our understanding of the semantics of the verb includes information about possible relationships in a sentence--for example, whether an agent (or an object, an instrument, etc.) can be present.³ If checking the verb--and to some extent, the heads of other phrases--turns up a set of relations that is incompatible with the first syntactic analysis, then the interpreter must construct a new analysis. But if the set of relations is consistent with the syntactic analysis, the sentence is easily processed.

A cross-linguistic study by Bates, McNew, MacWhinney, Devescovi, and Smith (1982) provides a different kind of evidence that an English speaker's first strategy in interpreting a sentence is to map an SVO structure onto it, despite conflicting semantic information. Bates et al. describe preferred syntactic mapping in terms of *prototypes* that may contain a number of other notions. For example, the prototype surface grammatical category "subject"⁴ may combine the functional notions

³These terms are used by Fillmore and others in describing a case grammar, but the point is simply that part of what native speakers know about verbs is what kinds of sentences the verbs can appear in.

⁴Surface grammatical categories correspond basically to familiar traditional categories such as subject and object. They are determined entirely by the order in which words occur in the sentence, not by semantic information. For example, in both of the following sentences, "governor" is the agent:

- (1) The governor asked the expert to testify at the hearing.
- (2) The expert was invited by the governor to testify at the hearing.

However, in the first sentence, "governor" is also the surface

What kinds of mapping strategies for sentences do English speakers use? Some strategies may be universal. For example, the overwhelming majority of the world's languages place subjects before objects (Greenberg, 1966; Pullum, 1977). Other strategies are language-specific. A substantial body of empirical research suggests that English speakers use the basic strategy of mapping a subject-verb-object (SVO) syntactic structure onto sentences they are attempting to interpret, even in the face of conflicting semantic information. Of course, when the preferred syntactic mapping does not agree with the semantic interpretation of the sentence, a new syntactic interpretation is made. But reading times are faster for those sentences in which the preferred structural analysis and the semantics match.

Two of the more compelling studies in this area are discussed below.

Rayner, Carlson, and Frazier (1983) conducted two experiments to explore the effects of semantic and pragmatic information on the syntactic analysis of ambiguous sentences. They recorded the eye movements of subjects reading structurally ambiguous sentences such as the following:

- (A) The maid passed the caviar didn't eat any of it.
- (B) The lawyer sued for damages lost the lawsuit due to a technicality.

The experiments showed that the relative plausibility of two possible real-world events does not influence the language processor's choice of an initial syntactic analysis for an ambiguous sentence, and that "reading times are longer when the most plausible analysis does not correspond to the analysis selected by the processor's structural preferences" (Rayner, Carlson, and Frazier, p. 371). For example, reading time was always longer for sentences such as

- (C) The performer sent the flowers was very pleased.

than for sentences such as

- (D) The performer sent the flowers and was very pleased with herself.

- Readers process actively as they move through a sentence, from left to right.
- Readers appear to process lexical, structural, semantic, and contextual information in parallel, using whatever information is available in a maximally efficient way (Marslen-Wilson and Tyler, 1980; Rayner, Carlson, and Frazier, 1983).
- Readers do not wait to experience the entire sentence before integrating these various sources of information to assign a meaning. They anticipate the structure and words to come and form a hypothesis (interpretation) about the rest of the sentence.
- If, as they proceed through the sentence, readers discover that their initial hypothesis was wrong, they use all available information to diagnose the source of the error and selectively reanalyze the sentence, focusing on that part of the initial analysis that caused the problem (Frazier and Rayner, 1982).
- Languages and syntax take certain forms because humans process information in certain ways; thus, language processing does not differ in *kind* from other cognitive abilities.

Humans have limited cognitive resources, and in many activities they call upon automatic processing strategies in order to use these resources efficiently. Automatic processing allows cognitive resources to be allocated to more difficult--that is, less predictable--demands. Even higher-level linguistic processing appears to have an automatic component (Britton, Glynn, Meyer, and Penland, 1982; Bock, 1982). On the sentence level, this automatizing consists of mapping strategies for interpreting sentence syntax. To the degree that the actual sentence and the mapping strategy coincide, resources can be committed to more difficult aspects of interpretation. Language acquisition provides some evidence that speakers do allocate cognitive resources to the most difficult tasks. Children rely very strongly on regular word orders, even when the language they are acquiring has a fairly variable word order (Braine, 1976), and they choose simpler syntactic structures when lexical content is more difficult (Bloom, Miller, and Hood, 1975).

The CLARIFY system provides a working-model answer to these requirements. The system and the rationale for designing a sentence-level revision guide are described in detail in the following section.

THE ROLE OF THE SENTENCE IN UNDERSTANDING TEXT

Although many important components of readability must be described at the text level--indeed, many psychologists concerned with readability have stopped looking at grammatical complexity and have concentrated instead on the propositional complexity of a text--research strongly supports a continued concern for complexity at the sentence level. The sentence is an extremely important component of comprehensibility.

Several researchers (e.g, Kintsch, 1974; Romelhart, Lindsay, and Norman, 1972) have proposed a system for memory representation based on Fillmore's case grammar. In this model, the verb is central, specifying the semantic relationships that tie the other sentence components together. Readers use individual sentences to construct propositions, then put the propositions together to form macro-propositional structures for the whole text. Features such as the complexity of ideas or the number of inferences made affect the ease with which the reader can construct the macro-propositional structures. It appears that the easier it is to process individual sentences, the more readily the sentences are moved from the short-term memory buffer and integrated into a text-level structure in long-term memory (see Miller and Kintsch, 1980; Isakson, 1979; Wisher, 1976; Fletcher, 1981).

We know a good deal about the strategies people use to process sentences.² Models of sentence processing developed in the last five years have shifted away from *transformational* grammar, in which syntax is an autonomous component, to *functionalist* models, in which syntax is the product of semantic, contextual, and lexical factors that influence the processing of messages underlying sentences.

The basic tenets of this new view of sentence processing can be summarized as follows:

²See, for example, Levin and Kaplan, 1971; Holmes, 1979; Isakson, 1979; Aaronson and Shapiro, 1977; Aaronson, 1976; Bock, 1982; Bates, McNew, MacWhinney, Devescovi, and Smith, 1982).

At Rand, as in many other corporations, a substantial amount of writing is never edited--interim research reports, corporate memoranda, brochures, progress and trip reports, proposals, etc. Much of this material is prepared on-line and could benefit substantially from the application of an on-line writing aid.

The needs of Rand's authors and audiences dictate certain requirements for a successful on-line revision guide. Such a guide must be

- Cost-effective in terms of both computer time and researcher time.
- Easily adapted to different audiences and situations.
- Usable by different kinds of people in different modes.
- Integrated with other corporate efforts to improve communication.

To be cost-effective, a writing aid must make good use of valuable researcher and computer time. Authors should not be forced to interpret statistical descriptions of text or to select from a laundry list features that indicate the need for revision. Instead of providing a global *ex post facto* assessment of the difficulty of text, the revision guide should focus the author's attention on those characteristics of sentences that relevant research suggests actually cause difficulty.

A successful revision guide must accommodate both a wide range of audiences and any special linguistic requirements of the authors' disciplines. Moreover, it should be usable both on-line and with hard copy, and at any stage in the composing process. To be maximally effective, the guide should be integrated with other corporate efforts to improve writing.¹

¹Rand provides a variety of in-house writing seminars for its research staff. They are taught by professional writing teachers and document designers. The seminars are described in Constance U. Greaser, *Improving Scholarly Writing at Rand*, The Rand Corporation, P-6274-1, May 1979.

III. THE CLARIFY SYSTEM: RATIONALE AND DESIGN SPECIFICATIONS

DESIGNING AN ON-LINE REVISION GUIDE

Despite the uncertain prospects for existing readability formulas, the mushrooming growth of automated text processing makes it tempting to try to harness the computer's speed and flexibility to the task of assisting in text revision. CLARIFY is the result of a research project undertaken at The Rand Corporation to use the features of a text-editing system to prompt an author to revise prose. The environment in which it was developed is representative of many organizations and businesses in which scholarly and scientific writing is an important corporate product.

Rand documents exhibit several general features of scientific and scholarly writing:

- They have varied, sophisticated audiences that can be generally described as civilian and military decisionmakers and technocrats.
- Difficulty of vocabulary is not an issue for these audiences. although jargon may be.
- The corporation has no contractual requirements to meet any designated reading level.
- The documents are not intended to be classroom texts.

The way in which Rand documents are written, edited, and produced is probably also widely representative. Some authors draft and revise entire manuscripts on-line, then send computer files to editors who edit and code them for phototypesetting. Other authors compose drafts on the typewriter or in longhand, then give their manuscripts to a text-processing specialist or secretary to enter into the computer for subsequent revision and production. More than 80 percent of all Rand documents become computer files during the production phase, and the number is growing.

global, and because the measures are not causal, the formulas do not identify those sentences that may need revision.

Using readability measures to guide revision can have perverse consequences. Several studies of the effects of using readability formulas to guide text revision have shown that while revisions did indeed improve the readability score, the changes generally had no effect on comprehensibility (Bruce, Rubin, and Starr, 1981). Indeed, revisions "according to formula" actually increased the difficulty of the text (Davison, Kantor, Hannah, Hermon, Lutz, and Salzillo, 1980; Davison and Kantor, 1982; Charrow and Charrow, 1979). For example, shortening sentences resulted in better readability scores, but it also eliminated explicit signs of relationships between propositions such as *consequently*, *however*, *in addition*, etc., which provide important cues to the reader. Other sentence-shortening devices, such as deleting relative pronouns or the complementizer *that*, make sentences more difficult for readers to process because they omit important cues to syntactic structure (Fodor and Garrett, 1967). Because readability formulas may make individual sentences more difficult to understand--to say nothing of the text-level features of comprehensibility that they ignore--they should not be used as guides to writing.

Readability formulas probably should not be used even for global predictions of text difficulty if the current readers are different from the readers with whom the formula was validated. This is particularly true when the groups differ in cultural background, dialect, or age and sophistication. No algorithm should be applied to a population that differs significantly from the one on which it is based. The aforementioned McCall-Crabbs Standard Test Lessons, which are widely and inappropriately generalized from their original population of New York City elementary school children to other populations, are a striking example of such poor statistical practice.

The second major limitation of readability formulas is that they are not particularly good measures of readability. In the last 15 years, research in linguistics, psychology, reading, and other fields has shown that it is neither useful nor appropriate to define readability as a set of superficial text features to be captured by the correct algorithm. Readability is in fact a complex interaction among features of text and the processing strategies and resources of readers (Miller and Kintsch, 1980).

Readability defined in this way has little or no connection with readability formulas. For example, word frequency and sentence length affect reading time but have little effect on memory (Miller and Kintsch, 1980). Flesch scores are independent of recall and are therefore poor indicators of readability. There appears to be no relationship between Fry readability formulas and analysis of text difficulty based on a text grammar (Templeton, Cain, and Miller, 1981). Surface difficulty measured by readability formulas does not correlate with difficulty in understanding and retaining information. In a series of experiments, Duffy and Kabance (1982) evaluated the effects of applying readability guides to text revision by simplifying both vocabulary and sentence structure and testing subjects' ability to perform tasks and to learn material after reading the modified texts. With one exception, these manipulations had no effect on comprehension, regardless of the skill of the participants.⁵

Because readability formulas are not good measures of text difficulty, it follows that they cannot guide an author in writing more readable text. The formulas provide general scores, averaged out over all the sentences in a tested sample. But because the assessment is

⁵Duffy and Kabance recommend management of the text production process through a *transformer*, an individual or group that ensures that a document is suitable for its intended purpose and audience. This approach to improving readability by influencing the entire composing and production process is similar to the approach developed at the Document Design Center, American Institutes for Research. At The Rand Corporation, professional writers and document designers work with researchers throughout the research and writing process. See *Simply Stated*, February-March 1982, and *IEEE Professional Communication Newsletter*, October 1982.

formulas can be summarized as follows:

- Readability formulas are not causal measures of text difficulty.
- They are not particularly good indicators of readability.
- They do not provide guidance for text revision.
- They do not have a generalizable database.

Readability formulas are not, and were never designed to be, causal measures of text difficulty. The studies on which they are based were atheoretical--whatever worked was adopted. The criteria for success were speed and simplicity, two features that make these formulas appear particularly attractive for computer applications.

The formulas do have some predictive value, but they lack strong statistical support. Many formulas have been validated only against earlier formulas which, in turn, were validated against such classic tests as the McCall-Crabbs Standard Test Lessons in Reading (Bruce, Rubin, and Starr, 1981). However, the Test Lessons were not based on extensive testing, and the scores they yield lack comparability and reliability. The Test Lessons scores were derived from a limited population--children in grades 3 through 6 (in some cases, only grades 3, 5, and 6) in the New York City public schools. The Test Lessons were designed to be practice exercises; they were never intended to be used as a criterion for readability formulas (Stevens, 1980). Nor were they intended to serve as general indicators of reading ability across age, class, or cultural groups. The grade-level scores were "rough equivalents, provided for students to track their progress" (Stevens, p. 414). Nevertheless, the McCall-Crabbs Test Lessons remain the criterion for the Lorge, Flesch, and Dale-Chall formulas, as well as for many other later formulas.

Later validation studies are not much more reassuring (Klare, 1976). Only 39 of 65 studies showed a positive correlation between estimates of difficulty based on readability formulas and reader performance based on speed or comprehension; indeed, when comprehension is the variable being measured, only half of the studies show positive correlations with the predictions of readability formulas.

Another sophisticated computerized application of readability formulas is IBM's experimental EPISTLE project.³ The eventual goal of EPISTLE is to be able to fully parse business English--that is, to identify the part of speech each word represents and to specify its relationship to other words in the sentence. Ultimately, the system is intended to critique written material on points of grammar and style. In its present form, EPISTLE checks spelling and diagnoses five classes of grammatical errors: subject-verb agreement, wrong pronoun case, noun-modifier disagreement, nonstandard verb forms, and nonparallel structures. It also provides several levels of style critiques: word- and phrase-level critiques similar to those provided by the DICTION program; sentence-level critiques (e.g., "sentence too long," "too much distance between subject and verb"); paragraph-level critiques ("too many passive sentences," "too many compound or complex sentences," "poor readability score" (as measured by some standard readability formula)).

These critiques are a mixture of traditional readability measures and measures that research suggests cause difficulty in reading a sentence (such as distance between subject and verb). Each style critique has thresholds against which to compare its value. These thresholds can be adjusted to tailor the style critiques to individual environments.

The EPISTLE project is scheduled to be completed in five years. If the project produces an accurate parser of English, then EPISTLE could directly identify causal measures of text difficulty, rather than the surrogate measures that CLARIFY uses.

WHY READABILITY FORMULAS FAIL

Many researchers have discussed the imprecision of the global assessments derived from readability formulas and the limitations of these formulas as aids to text revision.⁴ The basic shortcomings of the

³See G. E. Heidron, K. Jensen, L. A. Miller, R. J. Byrd, and M. S. Chodorow, "The EPISTLE Text-Critiquing System," *IBM System Journal*, Vol. 21, No. 3, 1982; and *Business Technology*, July 1983.

⁴See, for example, Klare, 1974, 1981; Redish, 1980; Frase, 1981; Bruce, Rubin, and Starr, 1981; Davison, Lutz, and Roalef, 1981; Coke, 1982; Pearson, 1974; Selzer, 1983; Huckin, 1983.

Sentence Structure

Passives

This text contains a much higher percentage of passive verbs (44%) than is common in good documents of this type (22%). A sentence is in the passive voice when its grammatical subject is the receiver of the action.

Passive: The ball was hit by the boy.

When the doer of the action in a sentence is the subject, the sentence is in the active voice.

Active: The boy hit the ball.

The passive voice is sometimes needed

1. to emphasize the object of the sentence,
2. to vary the rhythm of the text, or
3. to avoid naming an unimportant actor.

Example: The appropriations were approved.

Although passive sentences are sometimes needed, psychological research has shown that they are harder to comprehend than active sentences. Because of this, you should transform as many of your passive to actives as possible. You can use the *style* program to find all your sentences with passive verbs in them by typing the following command when this program is finished

```
style -p filename
```

SOURCE: Macdonald, 1982.

Fig. 2 -- Sample output from the PROSE program

Those features of STYLE that provide calculations of readability indexes are now commercially available. More advanced features have not yet been released.²

²For example, the Workbench now has some programs that evaluate overall report organization. A program called ORG formats the text, preserving headings and paragraph boundaries, but it prints only the first and last sentences of each paragraph. The output allows authors to check topic and concluding sentences for each paragraph and may provide the structure for a good abstract.

readability grades:	(Kincaid) 12.3 (auto) 12.8 (Coleman-Liau) 11.8 (Flesch) 13.5 (46.3)
sentence info:	no. sent 335 no. wds 7419 av sent leng 22.1 av word leng 4.91 no. questions 0 no. imperative 0 no. nonfunc wds 4362 58.8% av leng 6.38 short sent (<17) 35% (118) long sent (>32) 16% (55) longest sent 82 wds at sent 174; shortest sent 1 wds at sent 117.
sentence types:	simple 34% (114) complex 32% (108) compound 12% (41) compound-complex 21% (72)
word usage:	verb types as % of total verbs to be 45% (373) aux 16% (133) inf 14% (114) passives as % of non-inf verbs 20% (144) types at % of total prep 10.8% (804) conj 3.5% (262) adv 4.8% (354) noun 26.7% (1983) adj 18.7% (1388) pron 5.3% (393) nominalization 2% (155)
sentence beginnings	subject opener: noun (63) pron (43) pos (0) adj (58) art (62) tot 67% prep 12% (39) adv 9% (31) verb 0% (1) sub conj 6% (20) conj 1% (5) expletives 4% (13)

Fig. 1 -- Sample tabular output from the STYLE program

PROSE provides several sets of standards for comparison, and authors may select the set of standards that will be applied to their text. The standards were developed from Bell Laboratories documents that were judged to be good; however, the criteria used have not been experimentally validated.

Another Writer's Workbench program is REWRITE. This program implements Lanham's theory (1979) that to revise prose, one should locate all prepositions, all forms of the verb *to be*, and all wordy phrases. REWRITE capitalizes all the prepositions and forms of *to be* identified by the STYLE program and capitalizes common wordy phrases that appear on a list specified by the program. The hard-copy output is intended to make potentially bad sentences visually obvious.

tasks. Aaronson (1976) found that for memory tasks, coding time increases as the reader moves through the sentence, and the primary focus is on the surface structure. However, in comprehension tasks, coding focuses on the subject noun, the verb, and the object, and on the relationships among them. Coding time decreases as the reader moves through the sentence because linguistic predictability increases.

SELECTING SURROGATE FEATURES FOR CLARIFY

An ideal text-revision aid would analyze every sentence to see whether an animate agent is in the subject slot and an active verb is in the verb slot. Lacking these features, it would check for second-order strategies--for example, the topic-object in the subject slot and the agent in a "by" phrase--that would speed the process of interpreting the sentence. However, that would require the computer to understand natural language. A second-best approach would be to have the computer parse the sentence, using morphological rules, probabilistic structural rules, and a dictionary of some sort. However, writing a sufficiently accurate parser is a very time-consuming and expensive undertaking.

In designing CLARIFY, we hypothesized that we could identify a high percentage of sentences needing revision by specifying certain patterns of surface features that the computer could recognize.

Revising a sentence so that it uses an active verb is the most effective way to move the sentence toward the optimal mapping prototype of "agent-verb-object." Where do verbs "go" if they do not appear in the main verb slot? In English, especially in technical prose, actions that could function as the main verb are often transformed into "things" by one of English's many nominalizing suffixes. Thus, *discuss* becomes *discussion*, *require* becomes *requirement*, *perform* becomes *performance*, etc.⁶ When the verb has been turned into a noun, it can no longer govern the grammatical relationships in the sentence, e.g., it can no longer have an object. Thus, sentences that lack the grammatical glue of a good main verb string nouns together with prepositional phrases or

⁶Obviously, it is sometimes appropriate to turn actions into things, e.g., when it is the thing that is being discussed. We have

pile them up in compounds such as *employee coverage termination*, *water subsidy distortion elimination*, or *information enhancement actions*.

Integrating these grammatical facts with our experience in revising thousands of sentences in Rand's writing workshops and in technical and policy documents, we hypothesized that sentences with certain patterns of nominalizations, prepositional phrases, and filler verbs would pose difficulties for the mapping strategies of readers.

We specified an initial set of patterns or flags and tested them against a database of 100 sentences selected from documents written by participants in Rand's writing workshops. The sentences had been used as examples of poor sentences that required revision.⁷ They were drawn from all of Rand's disciplines and research programs. We did not expect our patterns to tag all the sentences--only a substantial percentage. Of course, some sentences had problems, such as faulty logic, that had nothing to do with the lack of a main verb.

Based on the initial tests, we revised and expanded our patterns and tested them against another sentence database until they were flagging more than 80 percent of the sentences. We then tested the flags against random samples of text submitted to the Rand Publications Department for production. The resulting flagged sentences were checked against those marked for revision by one of the writing teachers.

The following flags were developed from these procedures:⁸

- A. Two or more nominalizations and two or more prepositions.
(*Example: The OSD role is generally one of policy formulation, allocation of resources, overview of service programs, and coordination among the services.*)

attempted to accommodate this fact in CLARIFY by specifying exceptions to the nominalization flag. In addition, it appears that writers nominalize in an unconscious attempt to make their prose sound significant. Sociolinguistic studies consistently show that readers unconsciously consider nominalizations the mark of important prose (McNeill, 1966; Williams, 1978; Hake and Williams, 1981).

⁷Rand's writing teachers pool their material, so some of the sentences had been used as examples by all 4 teachers, and all sentences had been chosen as examples by at least 2 teachers.

⁸These are the flags being used in the current test version of CLARIFY. Some special features of the system allow us to vary the flag specifications for research purposes.

- B. One or more forms of the verb *to be* and two or more nominalizations. (*Example:* The problem of verification of such restrictions is recognized.)
- C. Four or more prepositions. (*Example:* More initial clarity and planning about the goals of the data collection effort by individual networks and by NCI might have obviated later problems in using the data.)
- D. One or more forms of the verb *to be* and three or more prepositions. (*Example:* To determine the effect of a service member's occupational specialty on his reenlistment behavior, the specialties in the 1976 DoD personnel survey needed to be classified according to characteristics which might explain differences in the behavior.)

It is important to note how these patterns differ from traditional readability formulas, including computerized variations. First, the occurrence of a passive construction, which contains a form of the verb *to be*, will not automatically be flagged. A passive sentence with the agent in a "by" phrase will be passed over by both Flags B and D. (For example, a sentence such as "These policies have been proposed by every administration since Harding's" will not be flagged.) Not every passive should be flagged as needing revision. Passives are often necessary to set up the interaction pattern between sentences by which the comment (the new material at the end of a sentence) becomes the topic (the subject) of the following sentence. (For example, "These policies have been proposed by every administration since Harding's. And every administration has failed to generate congressional support.") At the sentence level, we have seen that looking for an object in the subject slot and an agent in a "by" phrase is an important secondary mapping strategy for English speakers. Of course, some passives are inappropriate and the sentences in which they occur should be restructured as active SVO sentences. However, our preliminary testing showed that most of these sentences have other characteristics--e.g., too many prepositional phrases or too many nominalizations--that cause them to be tagged by Flags B or D.

Second, CLARIFY does not deal with individual words at all. There is no list of taboo words or phrases; word and phrase selection is often a matter of taste and can be much more effectively handled by professional editors if it is addressed at all. There is also no list of "difficult words." Any notion of reading level established for textbook materials would not be applicable to the adult, well-educated audience for whom Rand researchers and their counterparts in thousands of corporations write. In addition, recent studies suggest that difficulty of vocabulary is probably not an issue for most adult audiences. Processing vocabulary seems to require the same amount of cognitive capacity, whether the words are common or rare (Britton, Glynn, Meyer, and Penland, 1982). For most adult audiences, identifying the meaning of words appears to have become an automatic skill.

Third, the flags make no reference at all to sentence length, although length would correlate strongly with sentences that are flagged. This is not to say that length is unimportant. Frase and Fisher (1977) showed that readers rate sentences more than 20 words long as less efficient. However, the difficulty in understanding long sentences is not a matter of length alone. Where and how the length occurs is more important. For example, the distance between the subject and the verb is crucial because complexity after the main verb is much easier to process than complexity before it. A sentence that begins with an agent-subject immediately followed by an active verb may contain a long complement structure and be readily understood. And whether or not sentence components are clearly marked is more important than how long the components are. The following pair of sentences illustrates the point:

- (1) A model that allows uncertainty about the cost and availability of oil to be specifically incorporated into fuel-planning decisions by utility planners is described here.
- (2) Here we describe a model that allows utility planners to specifically incorporate uncertainty about the cost and availability of oil into their fuel-planning decisions.

It is, of course, no accident that the flagged sentences tend to be long. Strong verbs are the key to shorter, tighter sentences. Their absence is usually marked by nominalizations or by strings of prepositional phrases. But the intent of the flags is to identify sentences that will probably make mapping difficult, not sentences that are simply long.

The initial specifications for CLARIFY were based on and calibrated against Rand documents, but we are confident that our assumptions about its structure would be equally valid in other agencies and corporations whose staffs write for an audience similar to Rand's. We have looked at and rewritten sections of more than 50 documents from other research institutions, aerospace contractors, lawyers, and businesses, and we have discussed our characterizations of the "difficult" sentences with many colleagues who work in such firms. All the evidence suggests that Rand writing is representative of all writing that has a strong technical component and is directed toward multiple audiences that include technical colleagues and decisionmakers.

IV. USING THE CLARIFY SYSTEM

HOW CLARIFY WORKS

The CLARIFY program works with other text-processing programs such as editors, formatters, and spelling checkers. To use CLARIFY, an author exits from the computer file he or she is writing in and uses a simple command to send the file through CLARIFY. The file sent is usually the raw (i.e., unformatted) file; CLARIFY ignores all embedded formatting codes. The author may specify either an on-line or hard-copy output. CLARIFY processes the text and produces a new file in which sentences that meet the structural description of any of the flags are indicated with a marker and with the letter of the appropriate flag. The marker allows the file to be searched quickly for tagged sentences. In the hard copy, flagged sentences are printed in boldface type. In addition, those elements in the sentence that caused it to be flagged are marked and labeled. Speed depends to some extent on the system's load, but it is about 5 seconds for start-up plus .75 second per double-spaced page.¹

When the CLARIFY program has been executed, the computer prints summary statistics of the following kind:

SUMMARY STATISTICS (Total for document)

Prepositions	:	619
Forms of "to be"	:	154
Nominalizations	:	184
Sentences	:	266
Flag A	:	38
Flag B	:	31
Flag C	:	70
Flag D	:	61
Flagged units	:	101
Not flagged	:	165

¹At Rand, CLARIFY has been implemented under UNIX Version 7 and under UNIX Berkeley Version 4.1.

The author may either print the flagged file or edit it on-line. Figure 3 shows a flagged file as it appears on-line, and Fig. 4 shows a hard-copy version of the same file. In the edit mode, the sentence initial marker is used as a search point so that the computer can move quickly from one flagged sentence to the next. At each flagged sentence, the author may revise the text or simply move on to the next flagged sentence.²

```
.pp
The problem we discuss in this section has two elements.
The first is to find a method for estimating the
relationship between spares investment level
and a direct, meaningful measure of system performance,
such as expected launch delay. [A][C] An explicit
[N]>representation [P]>of that relationship would allow us
to determine the spares investment level required to
support any specified level [P]>of system [N]>performance, or,
conversely, to specify the desired level [P]>of
[N]>performance [P]>in full light [P]>of its costs.
.pp
[D] The second element [P]>of the problem [V]>is to ensure that,
[P]>for each level [P]>of performance, the required
spares investment level [V]>is minimal. Each of these
components of the problem implies the other. [A][B][C][D] What
[V]>is needed, then, [V]>is not only an explicit [N]>representation
[P]>of the relationship [P]>between [N]>performance and cost,
but one such that each [P]>of its points [V]>is an optimum
[P]>in the sense that it represents the least-cost mix
[P]>of spares [P]>for its specific level [P]>of [N]>performance,
and, conversely, represents the best possible
[N]>performance [P]>for its specific level [P]>of investment.
.pp
[A][C] The [N]>computation [P]>of such a relationship depends [P]>on
estimates [P]>of component characteristics that emerge
[P]>from [N]>definition [P]>of the system's maintenance
concept and repair level decisions, and the
quality [P]>of the estimated relationship depends
[P]>on the quality [P]>of the estimates [P]>of component
characteristics.
```

Fig. 3 -- On-line version of text that has been flagged by CLARIFY

At the end of the revision session, the author exits from the file and removes all remaining flags with a single command. To check the

²The computer's search capability prompts the author by moving from one flagged sentence to the next. However, all users commented that they were led to revise other, unflagged sentences in order to make those sentences fit with flagged sentences that had been rewritten.

1 .pp

2 The problem we discuss in this section has two elements. The first is to
3 find a method for estimating the relationship between spares investment
4 level and a direct, meaningful measure of system performance, such as
5 expected launch delay. ■AC■ An explicit representation of that relationship
6 would allow us to determine the spares investment level required to support
7 any specified level of system performance, or, conversely, to specify the
8 desired level of performance in full light of its costs.

9 .pp

10 ■D■ The second element of the problem is to ensure that, for each level of
11 performance, the required spares investment level is minimal. Each of
12 these components of the problem implies the other. ■ABCD■ What is needed, then,
13 is not only an explicit representation of the relationship between
14 performance and cost, but one such that each of its points is an optimum in
15 the sense that it represents the least-cost mix of spares for its specific
16 level of performance, and, conversely, represents the best possible
17 performance for its specific level of investment.

18 .pp

19 ■AC■ The computation of such a relationship depends on estimates of component
20 characteristics that emerge from definition of the system's maintenance
21 concept and repair level decisions, and the quality of the estimated
22 relationship depends on the quality of the estimates of component
23 characteristics.

24

Fig. 4 -- Hard copy of text that has been flagged by the CLARIFY program

efficacy of the revisions, the author may send the unflagged file through CLARIFY again.

CLARIFY indicates where revision is needed, but it does not show how to do the revision. Its effective use assumes that the user knows the principles of revising sentences and needs only to be prompted by the flags and labels. Currently, these principles are taught in a special orientation session for new CLARIFY users. In addition, the system provides some revision reminders. A user revising on-line who is puzzled about why a sentence is tagged may ask for help by typing "help" and the letter of the flag in question. In response, the system provides the kind of assistance shown in Fig. 5. The B in front of the sentence indicates the kind of flag; the sentence-internal markers show nominalizations (N) and occurrences of the verb *to be* (V). Hard-copy versions of these help files are also available as part of the system's user documentation.

Some researchers choose to revise hard copy--that is, a printout of the flagged file. The line numbers on the printout correspond to the line numbers in the original file.

The hard-copy CLARIFY output makes the program available to authors who do not use a text-processing system. As long as the text becomes a computer file at some point in the production process, it can be run through CLARIFY. Anyone who types original material into the text-processing system may use CLARIFY.

What CLARIFY Identifies

CLARIFY flags sentences by using a combination of morphological rules, lookup tables, a verb dictionary, and lists of exceptions. It identifies prepositions and forms of the verb *to be* by simple lookup. It uses the verb dictionary to distinguish infinitive forms (e.g., to deter) from prepositional phrases beginning with *to* (e.g., to airports). It also distinguishes between participial forms beginning with *by*, *for*, etc. (e.g., for maintaining, by attacking) from prepositional phrases beginning with those same words. It does not count forms of *to be* that occur in progressive aspect verbs (are breaking, were continuing). It scores "units" rather than sentences. Ends of units are marked by the

FLAG B: One or more forms of "to be" and two or more nominalizations

This sentence was flagged because it makes excessive use of the verb "to be". The sentence also contains many nominalizations (verbs made into nouns) strung together with prepositional phrases. To improve the sentence, replace forms of "to be" with active verbs. Whenever possible, make the agent of one of those verbs (i.e., the doer of the verb's action) the subject of the sentence.

The following sentence is an example of the same kind of problem. The revision shows how to fix it.

Example:

■B■ The problem of ■N>verification of such ■N>restrictions
■V>is recognized but it ■V>is either presumed an answer will ■V>be
found or it ■V>is argued ■N>restrictions ■V>are necessary and we
simply must accept the chance of cheating.

Revision: (new verbs capitalized)

We RECOGNIZE the problem of verifying such restrictions; however, we
PRESUME either that we WILL FIND an answer or that we simply MUST
ACCEPT the chance of cheating.

Fig. 5--CLARIFY response to a user's request for assistance

occurrence of the punctuation marks .?! or ;. Thus, compound sentences joined by a semicolon are scored as separate units; those joined by a comma and a coordinating conjunction are scored as single units. CLARIFY ignores all material enclosed within quotation marks.

One of the most important features in CLARIFY is the use of exceptions. Some of these reflect the fact that phrases such as *for example*, *on the other hand*, or *in addition*, have the form of prepositional phrases but the function of discourse markers that connect sentences. They are very desirable features of text. Other exceptions reflect the environment in which CLARIFY operates. There are many words that are nominalizations in form, but for which a verb cannot usually be satisfactorily substituted in technical writing. Examples include *assistance*, *attrition*, *communication*, *consumption*, *desegregation*, *distribution*, *gestation*, *information*, *litigation*, *motivation*, *organization*, *production*, *relation*, *settlement*, *transportation*, *treatment*, and *variance*. An exception list keeps CLARIFY from identifying such words as nominalizations and thus tagging sentences for revision that cannot be satisfactorily revised by eliminating the

nominalization. The current version of CLARIFY has a general list that is appropriate for all Rand programs. It would be possible to tailor the exceptions to individual research programs, and we are studying that possibility. However, it is desirable to keep the exception list as short as possible, since authors are always tempted to assume that all their own nominalizations are necessary.

Detailed Description of the Structure of CLARIFY

CLARIFY is not really one program, but three. They appear as one program to the user, however, because all three are invoked at once by means of another single program. Each of the three CLARIFY programs has one input and one output. They feed into each other as follows:

INPUT-FILE => PHRASE => POS => FLAG => OUTPUT-FILE

We discuss each of the three programs in turn.

PHRASE is a lexical analyzer that identifies exception phrases (e.g., *in conclusion*, which is not to be identified as a prepositional phrase) and certain other features, including the end-of-sentence and some common abbreviations. These identified features become "tokens." When the text flows out of PHRASE, some of it has been classified as "tokens" and some of it has not yet been classified.

The second program is POS, which stands for part of speech. It is a lexical analyzer that classifies words into "tokens" representing parts of speech, namely, nominalizations, prepositions, and forms of the verb *to be*. Other words are classified only in order to assist in later processing, and many words are classified as "unknown." POS uses a combination of word endings and a small dictionary. Obviously, it will produce some incorrect classifications, but it works well enough for the sentence-level characteristics we are looking for.

The final program, FLAG, determines whether or not to flag a given sentence. It uses the input tokens identified by the previous two programs and assigns the four sentence flags, as appropriate. FLAG also keeps summary statistics about the document.

INITIAL TESTS OF CLARIFY

Development of CLARIFY began in the fall of 1979. A preliminary system was available in early 1981 for user tests with eight members of the Rand research staff representing a variety of research interests and backgrounds and different levels of computer sophistication. All of the test users were accustomed to drafting and revising text on-line, and they shared a concern for producing clear concise documents and for controlling the costs of editing and production.

In an orientation session, the users were given instructions about how to use CLARIFY, as well as written documentation. They were given written guidelines for evaluating the system and were asked to keep notes as they used CLARIFY. (The guidelines are reproduced in Appendix A.) The users were urged to document their immediate reactions after first use and to provide a hard-copy printout of all the versions of each file that they processed through CLARIFY. We then used these materials to investigate how many iterations of CLARIFY appeared to be most cost-effective.

After two months, we personally interviewed each user. (The questionnaire we used in these interviews is reproduced in Appendix B.) The users' comments are summarized below:

- All of the users said that they were initially very skeptical when they saw how many sentences CLARIFY had flagged in their text. However, they also all agreed that once they examined the sentences *from the perspective of being "told" that the sentences needed revision*, they decided that rewriting was really necessary.
- All said that because CLARIFY focused attention on specific sentences, revision was more efficient, less painful--"worth two rereadings." Several commented that the prospect of revising an entire text was extremely daunting. They found it very helpful to have CLARIFY direct their revision to specific sentences and specific aspects of those sentences.

- All found that revision became easier as they moved through the flagged text.
- All said that they initially ignored the unflagged sentences, but found that they often revised them later because of the changes in the flagged ones.
- All said that a surprise benefit was the teaching power of the system: They felt that after using CLARIFY several times and revising under its direction, they began to "internalize" the flags and produce better prose to start with.
- All users reported revising at least two-thirds of the flagged sentences.
- All but one user preferred to revise a hard-copy version of the flagged file, then insert the revisions into the original file or have them inserted by someone else.

Inevitably, the users differed in their reactions to the system's details. One user thought the flag labels were not very useful: If a sentence was flagged, he simply went back and "rethought" the sentence. Another found the nominalization flags of little help but the preposition flags very useful. Several others considered the labels and the pointers inside the sentences to be key to the revision process, but one expressed concern that in revising to avoid preposition flags, he was tempted to create strings of nominal compounds.

More important were the differences in users' attitudes about the appropriate interactions among the system, the revision process, and the editing process. As a matter of corporate policy, The Rand Corporation requires all official research documents to be edited. Because editing costs are charged against the research budgets, researchers are constantly looking for the best tradeoff between spending more of their own time revising and asking editors to do more work on a manuscript. Thus, if CLARIFY is to have widespread use, researchers must feel that it is cost-effective for them to revise, using CLARIFY, rather than to leave all revision to an editor.³ Most of the researchers felt that

³Our experience revising thousands of sentences during four years of Rand's writing workshops convinced us that the author should do most

they were revising much more thoroughly and efficiently when they used CLARIFY. Thus they felt that their documents would ultimately require lighter editing, and less time would be spent resolving changes of meaning that occur in the editing process. However, two researchers took diametrically opposed positions on this issue. The first, an economist, felt that an editor could make revisions more efficiently and effectively, but he was most enthusiastic about the system's teaching benefits. He also expressed the view that the program's "net" was too fine--i.e., that CLARIFY was tagging too many sentences.⁴ At the opposite extreme, the second researcher, a statistician, found using the system superior to most of his interactions with editors. Although he writes few complete documents, he frequently contributes technical discussions to larger pieces. He felt that editors often changed the meaning of his material because they didn't understand it. Reestablishing the correct meaning wasted his time, he felt, so he became impatient with the entire editing process and demanded that no changes be made. In contrast, CLARIFY provided him with an "objective" measure of his prose. He was pleased to have his attention focused on certain sentences (20 to 40 percent of the total) and claimed to have revised nearly all of them--more than twice as many as he thought he would have revised without a revision guide. Because he writes many technical memos that are not edited, he finds this increased revision efficiency the major benefit of CLARIFY.

One of Rand's editors, an experienced text-processing user, also agreed to be one of our initial users. We wanted her opinion about the potential utility of such a system for editorial purposes.

of the substantial sentence-level revision. In particular, we learned that the selection of the main verb for a sentence entailed at least a choice among emphases and at most a choice among meanings. That choice is best made by someone familiar with the substance of the text. When editors make these kinds of changes, they risk distorting meaning. Good editors are perfectly aware of this risk and often flag changes they have made so that authors will review them carefully in order to prevent possible distortions.

⁴None of the other experimental users complained that too many sentences were flagged.

(15) Did you consult the Help files? Were they convenient to get to? Did they give you the information you wanted?

(16) Did the orientation session adequately prepare you for using CLARIFY and for revising the sentences it tagged?

(17) What do you see as the biggest benefit of this system?

(18) What would you most like to see changed?

(19) What is your general assessment of the system:

- o useless because you would have revised the sentences anyway;
- o potentially useful but too much of a nuisance to use;
- o valuable guide for revising sentences

Other comments?

Appendix B

QUESTIONNAIRE USED TO EVALUATE CLARIFY IN INTERVIEW WITH TEST USERS

- (1) Did you read the general documentation of CLARIFY before you started to use it? Was the documentation clear?
- (2) Did you revise on line or on hard copy?
- (3) Once you used CLARIFY, could you easily see which sentences had been flagged when you were working on line? When you had a laser printout?
- (4) Can you easily interpret the labels used inside the sentences to indicate why the sentence was tagged? (If hard copy, could you easily interpret the change in font?)
- (5) Is it useful to focus on the elements that caused the sentence to be tagged or did you ignore the labels?
- (6) Did you find it easy to move through your flagged file using the Search key, or did the flags interrupt your normal revision process?
- (7) Roughly what percentage of the flagged sentences did you actually revise? How many of these would you have revised anyway?
- (8) Do you feel CLARIFY is tagging too many sentences?
- (9) Was it daunting the first time you saw how many sentences were tagged? Was it annoying to have the computer do this to your prose?
- (10) Do you feel you spent more time revising when you used CLARIFY than if you had been doing a routine revision?
- (11) Did you feel obliged to revise a sentence because it had been flagged?
- (12) Do you feel that the changes you made in response to CLARIFY could have been made easily and accurately by an editor?
- (13) Did you notice any change in the pattern of your revision as you moved through your file? Did it become easier?
- (14) Have you noticed any change in the way you write since you have been using CLARIFY?

Appendix A

SUGGESTED GUIDELINES FOR EVALUATING CLARIFY

We will ask you to give us two brief interviews about your reactions to CLARIFY: the first after you use the system for the first time, the second after you have more experience with it. As you use CLARIFY, please keep some informal notes based on the guidelines below:

- (1) Did you read the general documentation of CLARIFY before you started to use it? Was the documentation clear?
- (2) When you are working on line, can you easily see which sentences have been flagged? When you have a laser printout?
- (3) Can you readily interpret the labels used inside the sentences to indicate why the sentence was tagged?
- (4) Did you find it easy to move through your flagged file using the Search key, or did the flags interrupt your normal revision process?
- (5) Roughly what percentage of the flagged sentences did you actually revise? What percentage of the flagged sentences would have revised anyway?
- (6) Do you feel CLARIFY is tagging too many sentences?
- (7) Did you consult the Help files? Were they convenient to get to? Did they give you the information you wanted?
- (8) Did the orientation session adequately prepare you for using CLARIFY and for revising the sentences it tagged?
- (9) What is your general impression of the system:

- useless because you would have revised the sentences anyway;
- potentially useful but too much of a nuisance to use;
- valuable guide for revising sentences.

Whenever you use CLARIFY, please provide Mary Vaiana with

- (1) a hard copy of any file you send through CLARIFY
- (2) a laser copy of the output of CLARIFY
- (3) a hard copy of the revised file

These data will *not* tell us whether it is cost-effective--in any normal sense of that term--for a researcher to use CLARIFY. Indeed, because it is virtually impossible to obtain consistent ratings for any kind of writing, the ratings of each revision should be interpreted as merely suggestive. Authors themselves must decide when they should spend time revising and when they should rely more heavily on editorial assistance. But we will gather some information that can help inform that choice. For example, we will know, on average, how long it takes authors of various writing abilities to revise, and how that compares with editorial time. Other factors that authors might consider in deciding who should revise include the kind of document, the kind of audience, the extent to which CLARIFY complements their usual revision procedures, and their current workload. In particular, these data should indicate how valuable CLARIFY is for documents that will not be edited, and they may suggest guidelines for the system's routine use on them.

CLARIFY has also been integrated with other corporate efforts to improve writing. We made CLARIFY available to a group of researchers who participated in a writing workshop in March 1983. The workshop's main emphasis was on strategies for organizing documents effectively. However, the instructor spent one 2-hour session explaining techniques for revising sentences and practicing these techniques with the workshop participants. She then explained CLARIFY as a simple way to reenforce those techniques. Although using CLARIFY in this way constitutes a separate experiment, we are collecting the same data from these researchers and from the 20 participants in the test group, and we are also collecting the same information on composing styles.

ranking provided us with both a crude baseline assessment of each researcher's writing ability and an opportunity to select researchers with a range of writing skills.

All the members of the test group will be asked to fill out a composing-styles questionnaire (Appendix C). In addition to providing information about writers' composing and revision habits, the questionnaire will enable us to check the validity of some of the observations of the first test users. For example, the questionnaire asks researchers to describe their perception of a typical interaction with an editor. We will also collect a 10-page writing sample from each test user on which we will base an evaluation of the system's "teaching effect."

EVALUATION PROCEDURES

The questionnaire on composing styles will enable us to answer many questions about who uses CLARIFY. We will tap user satisfaction through an interview with each user in the test group. This interview will solicit information about the user's *perception* of what it is like to use CLARIFY and revise under its direction. We are interested in the system's perceived effect on an author's willingness to revise and on his efficiency in doing so; the effect of this guided revision on the quality of the resulting drafts; and the comparisons between CLARIFY revisions and typical editorial interactions.

The efficiency and quality issues are intertwined. We need to determine whether our economist user was right in suggesting that an editor could make all the CLARIFY-indicated changes more efficiently. To investigate this issue, we will run a number of experiments like the blind-reading experiment described earlier. We will ask a subset of our test group to keep track of the time they spend revising a piece of prose with CLARIFY. The subset should represent a range of writing abilities. We will ask an editor to revise the same piece of prose and keep track of the time required. We will then get independent judgments of the quality of each revision from three professional writers, using some scale of "communication effectiveness." We will also ask the author to review the editor's version for accuracy.

Item	Public Users	Test Group
Data collection	Initial questionnaire. Short follow-up to measure usage.	Initial questionnaire. Interview. Time/effectiveness studies.
Selection criteria	None.	Vary writing ability.
Issues addressed	Simple satisfaction. Amount of usage. Usage for documents that are not edited.	Satisfaction (detailed). Amount of usage. Usage patterns. Revision time required. Quality produced. Usage for documents that are not edited.

Public Users. Whenever possible, we will ask new users to fill out a questionnaire on composing styles (see Appendix C). However, CLARIFY is used like any other program on Rand's text processing systems, so we will not necessarily know when a new person tries it. We will attempt to gather some very simple indications of user satisfaction and rate of usage by sending a questionnaire to all text-processor users after the program has been available for six months.

Test Group. The special CLARIFY test group consists of 20 Rand researchers who have accounts on the corporate text-processing system. To select them, we sent an initial screening questionnaire to all 150 users who have text-processing accounts, explaining the purpose of CLARIFY and asking the users if they would be willing to participate in a test program. Sixty-two were willing to test CLARIFY. We then asked the staff of technical editors to rank these researchers in terms of the levels of edit their documents ordinarily require.¹ This editorial

¹Rand editors are assigned to specific research programs. Since the members of the research staff tend to work on only one or two programs, the editors are usually quite familiar with the editing needs of each researcher in their program areas. Although there is inevitably some variation among editors, the levels of edit can be loosely defined as follows: light or copy edit (spelling, punctuation, grammar, and cursory check of figures, tables, and references); regular edit (copy edit, plus changes at sentence and paragraph level to maintain parallelism, make structures more understandable, add transitions); heavy edit (substantial rewriting).

How Well Will Users Like CLARIFY?

6. Will most users feel that focused revision saves them revision time?
7. Will most users feel that CLARIFY prompts them to do more extensive revision than they would have done on their own?
8. Will expectations about editing influence the evaluation of CLARIFY? For example, will authors who rely on an editor to do heavy rewriting find CLARIFY less useful?
9. Will authors who feel that editors often change meaning feel that CLARIFY is more efficient than working with an editor?
10. Will writing ability influence the evaluation? For example, will writers who usually require only light editing find CLARIFY a useful guide to revision, while those who usually require heavy editing find the system annoying or too strict?

How Will CLARIFY Affect Authors?

11. Will users feel that revision becomes easier as they move through a document?
12. Will authors who use CLARIFY regularly experience a "teaching effect"?

What is the Best Way to Use CLARIFY?

13. Will the first pass through a text with CLARIFY remain the most effective?

Does CLARIFY Produce Revisions of Acceptable Quality?

14. Will independent assessments of text revised by writers of differing abilities using CLARIFY suggest that the revised text is comparable in quality to that produced by a human editor?

USER GROUPS

The chart below summarizes the ways in which we intend to evaluate the use of CLARIFY with two groups of users:

V. CURRENT TESTING PLANS

In November 1983, CLARIFY was made available for general use on Rand's text processing systems. While CLARIFY is in general use, we will continue to evaluate it in a number of ways, both with the entire user community and with a smaller subset of that group who have agreed to participate in a more structured test program. After this period of more extensive testing, CLARIFY is expected to be made available to the public in March 1984.

RESEARCH QUESTIONS

There are five categories of research questions that we hope to address during our tests of CLARIFY.

Who Will Use CLARIFY?

1. Will authors' use of CLARIFY be influenced by the typical audience for their documents?
2. Will authors' use of CLARIFY be influenced by the usual purpose of their documents?
3. Will writing style influence use of CLARIFY? For example, will authors who prepare a detailed outline and try to write a fairly polished first draft be less likely to use CLARIFY than those who just try to get material down in a first draft?
4. Will the method of drafting material influence the use of CLARIFY? For example, will authors who usually compose drafts on-line use CLARIFY more than those who usually compose drafts in longhand or on the typewriter?
5. Will the method of revision influence the use of CLARIFY? For example, will authors who usually revise on-line or on the hard copy of a computer file use CLARIFY more than those who revise a handwritten or typewritten manuscript?

One additional criterion that should be included is ease and flexibility of use. CLARIFY satisfies this criterion because it can be used with either hard-copy or on-line output; it can thus be used by authors who do not use a text-processing system and can be used as a revision guide at any time in the writing process from first draft to final editing.

By Frase's criteria, CLARIFY gets good, but not outstanding marks. Its most severe limitations are its sentence-level focus and its use of measures that are not directly causal. Its strongest features are its capability to guide revision, its potential for improving first drafts, and its adaptability to different user needs. Its real efficiencies and the most fruitful ways in which it can be used have yet to be clearly defined.

5. Provide explanations and interpretations along with measures.

CLARIFY provides an explanation in the form of labeled tags, a help file with a longer explanation, and an example to aid revision.

6. Make relative, not absolute, evaluations: Show how a text deviates from other texts selected as standards.

CLARIFY assumes that the reader's mapping strategy at the sentence level is the same for all texts, and therefore that good sentences in all texts would have many common features. However, the use of exceptions to the nominalization specification and the possibility of making those exceptions specific to each research area constitute a kind of relative evaluation.⁸

7. Allow users to set standards.

CLARIFY could be tailored to individual user groups. For example, we could build exception lists that are research-area-specific, or we could allow certain users to experiment with changing the flag specifications in order to widen or narrow the net.

8. Treat measures as information, not decisions.

CLARIFY makes no alterations to the text; it simply marks it in specific ways. Its successful use depends on the trained judgments of the author.

* * * * *

This is not the only set of first principles in the universe of readability research and document design, but it is a reasonable set.

⁸It would, of course, be possible to build a corporate database of texts, similar to that maintained for Writer's Workbench, against which authors could calibrate the CLARIFY statistics from their own texts.

research-area-specific lists of phrases that authors could invoke; however, the preposition issue must be tested with a wider group of users before we can make such arrangements.

ASSESSING CLARIFY

In the context of a warning to use new technologies wisely, Frase (1981) lists eight principles that guided the development of an automated editing system at Bell Laboratories. These measures and our attempts to calibrate CLARIFY against them are given below.

- 1. Build causal measures: Use measures that research indicates have consequences for text comprehension and use.⁷**

The measures used in CLARIFY are not directly causal. They are surrogates for measures that research clearly shows have consequences for text comprehension.

- 2. Build sophisticated measures: Include measures that may tap organization, such as ratios of nouns to verbs and formulas of the distance between repeated words.**

The current version of CLARIFY functions only on the sentence level, but the framework would make it possible to add some text-level features such as use of connectives and position of subordinate clauses.

- 3. Use multiple measures: Measure many variables to maintain a properly complex perspective.**

CLARIFY uses four measures--all at the sentence level.

- 4. Create measures that point to questionable aspects of a text.**

CLARIFY directs attention to questionable sentences and to the questionable elements within them.

⁷In the strictest sense, only a program that parses text can support truly causal measures.

If the author's revision is indeed that good, we can expect his future documents will be much easier for an editor to work on. Thus the editor can begin revision at a much higher level, and the resulting document should be better than the mere sum of the two revision efforts. It should also be less costly.

While we were developing the first test version of CLARIFY, we used a random sample of documents submitted by researchers for publication as well as material being drafted by professional writers in the corporation. We discovered that the pattern of flags corresponded to two basic categories of prose that we all recognized: dense, highly nominalized prose in which the reader must struggle to discover the relationship among stacks of nouns, and spaghetti-like prose in which the reader struggles to determine who did what to whom. Text that primarily triggered flags A and B fell into the first category; text triggering flags C and D fell into the second.

We also discovered that the professional writers very rarely wrote sentences that got A or B flags, but they did write sentences that got C and D flags; and some of these did not yield to effective revision. While it is not surprising that good English sentences can contain four prepositional phrases, our samples suggested an interesting hypothesis, which we intend to pursue: When four prepositional phrases are necessary in a sentence, they will usually be distinguished by function (time, place, manner, etc.) or they will be bound to the head noun in a special way, for example, partitives like "four of the group." Both distinctions in function and the restrictions that English places on the ordering of time, place, and manner elements provide good cues to the reader.⁶ And our initial users suggested to us that the system should find a way to handle "bound phrases" such as "value of children" or "economies of scale" that were really single cognitive units but would be counted as prepositional phrases. It would be possible to build

engineers and managers spent more than 30 percent of their time writing letters, reports, memos, and proposals. See *Technology Review*, April 1983.

⁶The normal order of occurrence for adverbial elements in a sentence is *manner*, *means*, then any order of *instrument*, *place*, and *time*.

We also attempted to compare the revisions done by authors with those done by editors. We asked two of Rand's writing instructors to evaluate two versions of a 5-page section of a document. The first was the author's own revision of his first draft, using CLARIFY. The second was the edited version produced by a senior research editor. The editor had been given a copy of the author's first draft and asked to edit it to his satisfaction. The instructors did not know anything about the two versions; they were simply asked to judge the effectiveness of the communication. In addition, we gave the two versions to one of the author's professional colleagues and asked her to evaluate their relative effectiveness.

The writing instructors had contrasting opinions. One thought that the author's version was better structured in terms of presenting the content but that the editor's version was more "graceful." The other found the editor's version superior. The author's colleague saw no difference in the versions and found both effective.

These inconclusive results are not surprising, given the small size of the sample and the well-known inconsistency that besets any evaluation of writing. Nevertheless, they suggest the following:

- Empirical tests of the system's effectiveness will be extremely difficult to construct.
- Matters of taste may make it impossible to compare the effectiveness of author revision and editorial revision. Even if it were possible to construct some usable definition of "grace of expression," we do not know of any studies of the effects of grace of expression on comprehension.
- If the author's revision is comparable in quality to the editor's, we can perhaps be optimistic about the role of CLARIFY in revising manuscripts that are not to be edited. If subsequent testing shows that CLARIFY can prompt an author to produce clearer prose without the benefit of an editor, a system like CLARIFY could be very useful in organizations where staff and managers spend substantial amounts of time writing.⁵

⁵A recent study of a division of Exxon Chemical Company by members of the Technical Communication Group at M.I.T. showed that staff

Like most of the researchers, she preferred working with hard copy. She thought the system flagged many sentences that she found acceptable; however, she noted that this was a matter of style. (The version of CLARIFY she used did not treat compound sentences as separate units, as the current version does, and her major complaint was that too many compound sentences were flagged.) She did think that the system provided her with a tool to calibrate her style periodically, and she felt the experience with CLARIFY had made her more aware of lapses into careless habits. For this reason, she thought she might like to use the system as a last pass over a document to check for editorial lapses. But she thought the system's real value would be for the authors: "A critic that can be invoked and used in private has a good chance of being heeded."

EVALUATING EFFECTIVENESS

The enthusiastic reception by our first users, along with a small experiment (described below), convinced us that CLARIFY was worth testing on a larger scale. The reports of the users and the text they were producing appeared to support our hypothesis that a system that identified *surrogate* features of text difficulty could prompt effective text revision. We were aware that finding ways to validate the system would be extremely difficult. In Sec. IV, we describe some of our approaches to that problem in the current test situation.

We were able to investigate the system's effectiveness very superficially with the limited data provided by our first test users. An important question from the point of view of cost-effectiveness concerned the number of times an author should run a file through CLARIFY. In several cases, researchers had used CLARIFY, revised the file, then resubmitted the file to CLARIFY. We had the summary statistics from these multiple runs--in one case, four runs; in another case, three. Both sets of statistics support our intuitive feeling that the most effective revision is done after one run; subsequent runs show very slight declines in the number of flagged sentences. The statistics are consistent with the researchers' feelings that revision is most fruitful the first time, and that subsequent revision is perhaps better left to an editor.

Appendix C
QUESTIONNAIRE ON COMPOSING STYLES

1. In which program do you do write most of your Rand documents?
(By Rand documents, we mean R's, N's, and P's.)

PROGRAM _____

2. What kinds of audiences do you write Rand documents for?

	OFTEN	SELDOM	NEVER	DON'T KNOW
A. Policy decisionmakers	1	2	3	9
B. Staff or advisory level	1	2	3	9
C. Research specialists	1	2	3	9
D. General audience	1	2	3	9
E. Mixed types of readers	1	2	3	9

3. What are the main purposes of your Rand documents?

	OFTEN	SELDOM	NEVER	DON'T KNOW
A. Document research methods and findings	1	2	3	9
B. Contribute to research in a field	1	2	3	9
C. Set context for policy debate	1	2	3	9
D. Evaluate policies	1	2	3	9
E. Explain research to non-technical audience	1	2	3	9
F. Propose new research	1	2	3	9

The next questions are about how you draft and revise Rand documents.

4. Which of these things do you usually do before you start writing?

	YES	NO
A. Think about the organization of the document	1	2
B. Write a sketch for the document	1	2
C. Prepare a detailed outline	1	2

5. When you write a first draft do you:

Just try to get it down on paper.....	1
Try to make it read fairly smoothly.....	2
Try to write a polished version.....	3

6. How do you prepare your first draft?

Handwritten.....	1
Typewriter.....	2
Text processor or Wylbur.....	3

7. How many drafts of a document do you usually produce before it is sent to a reviewer?

NUMBER OF DRAFTS_____

8. When you revise do you pay the most attention to:

Substance,.....	1
Style and organization, or....	2
Both about equally?.....	3

9. Do you usually revise:

On a handwritten draft,.....	1
On a clean, typed copy,.....	2
On a hardcopy of a computer file, or...	3
Directly on-line?.....	4

10. What level of editing comes closest to what you usually expect?

Copy edit only.....1

Complete edit.....2

Re-organization and rewrite...3

11. How much work do you find you have to do on a document once it is edited?

Extensive work to restore changed meanings and fix ambiguities.....1

Some work, but not extensive.....2

Little or no work before publication..3

12. Have you ever taken:

	YES	NO	PARTS OF IT
A. The Basic Effective Writing Course	1	2	3
B. The Advanced Effective Writing Workshop	1	2	3
C. The Briefing Workshop	1	2	3

13. Compared with other Rand writers, would you say your writing is:

Better than average.....1

About average.....2

Below average.....3

14. How heavy is your current and future writing load?

	IN-PROGRESS	IN NEXT YEAR
A. How many R's?	_____	_____
B. How many N's?	_____	_____
C. How many P's?	-----	-----

15. In the last year, how many proposals did you write all or part of?

PROPOSALS_____

16. In a typical month, how many memos do you write?

MEMOS_____

REFERENCES

- Aaronson, Doris, and Hollis Shapiro Scarborough. 1977. "Performance Theories for Sentence Coding: Some Quantitative Models." *Journal of Verbal Learning and Verbal Behavior*, 16: 277-303.
- Aaronson, Doris. 1976. "Performance Theories for Sentence Coding: Some Qualitative Observations." *Journal of Experimental Psychology: Human Perception and Performance*, 2, No. 1, 42-55.
- Aquino, Milagros R. (n.d.). Development of Linguistically-Based Readability Formulas. Study 5. Southwest Regional Educational Laboratory.
- Bates, E., and B. MacWhinnney. 1981. "Functionalist Approaches to Grammar." In L. Gleitman and E. Wanner (eds.), *Language Acquisition: The State of the Art*. New York, Cambridge University Press.
- Bates, Elizabeth, Sandra McNew, Brian MacWhinney, Antonella Devescovi, and Stan Smith. 1982. "Functional Constraints on Sentence Processing: A Cross-Linguistic Study." *Cognition*, 11: 245-299.
- Bever, Thomas G. 1970. "The Cognitive Basis for Linguistic Structures." In J. R. Hayes (ed.), *Cognition and the Development of Language*. New York, John Wiley and Sons.
- Bloom, L., P. Miller, and L. Hood. 1975. "Variation and Reduction as Aspects of Competence in Language Development." In A. Pick (ed.), *Minnesota Symposium on Child Psychology* (Vol. 9). Minneapolis, University of Minnesota Press.
- Bock, J. Kathryn. 1982. "Toward a Cognitive Psychology of Syntax: Information Processing Contributions to Sentence Formulations." *Psychological Review*, 89: 1-47.
- Bormuth, John R. 1969. *Development of Readability Analyses*. Final Report, Project No. 7-0052. USOE, Bureau of Research, U.S. Department of Health, Education, and Welfare.
- Botel, Morton, and Alvin Granowsky. 1972a. "A Formula for Measuring Syntactic Complexity: A Directional Effort." *Elementary English*, 49: 513-516.
- Botel, Morton, and Alvin Granowsky. 1972b. "A New Formula for Measuring Syntactic Complexity." Paper presented at International Reading Association Convention, Detroit, Michigan, April 1972.

- Braine, M.D.S. 1976. "Children's First Word Combinations." *Monographs of the Society for Research in Child Development*, 41: Serial No. 164.
- Britton, Bruce, Shawn M. Glynn, Bonnie J. F. Meyer, and M. J. Penland. 1982. "Effects of Text Structure on Use of Cognitive Capacity During Reading." *Journal of Educational Psychology*, 74: 51-61.
- Bruce, Bertram, Andee Rubin, and Kathleen Starr. (1981). *Why Readability Formulas Fail*. Reading Education Report 28. Center for the Study of Reading, University of Illinois, Urbana, Illinois.
- Cherry, Lorinda L. 1981. *Writing Tools--The STYLE and Diction Programs*. Computing Science Technical Report No. 91. Bell Laboratories.
- Cherry, Lorinda L. 1982. "Writing Tools." *IEEE Transactions on Communication*, No. 1, January 1982: 100-105.
- Charrow, R., and V. Charrow. 1979. "Making Legal Language Understandable: Psycholinguistic Study of Jury Instructions." *Columbia Law Review*, 79: 1306-1374.
- Coke, Esther U. 1982. "Computer Aids for Writing Text." In David H. Jonassen (ed.), *The Technology of Text: Principles for Structuring, Designing, and Displaying Text*. New Jersey: Educational Technology Publications, 383-400.
- Coke, Esther U., and Ernst A. Rothkopf. 1970. "Note on a Simple Algorithm for a Computer-Produced Reading Ease Score." *Journal of Applied Psychology*, 54: 208-210.
- Coleman, Edmund. B. 1968. "Experimental Studies of Readability. Part I: Stimulus Dimensions that Affect Readability." *Elementary English*, 45: 316-323.
- Dale, Edgar, and Jeanne S. Chall. 1948, "A Formula for Predicting Readability." *Educational Research Bulletin*, 27: 11-20, 37-54.
- Davison, A., R. N. Kantor, J. Hannah, G. Herman, R. Lutz, and R. Salzillo. 1980. *Limitation of Readability Formulas in Guiding Adaptations of Texts*. Technical Report 162. Center for the Study of Reading, University of Illinois, Urbana, Illinois.
- Davison, Alice, Richard Lutz, and Ann Roalef (eds.). 1981. *Test Readability: Proceedings of the March 1980 Conference*. Center for the Study of Reading Technical Report No. 213. Cambridge, Massachusetts, Bolt Beranek and Newman.
- Davison, Alice, and Robert N. Kantor. 1982. "On the Failure of Readability Formulas to Define Readable Texts: A Case Study from Adaptations." *Reading Research Quarterly*, 2: 187-209.

- Duffy, Thomas M., and Paula Kabance. 1982. "Testing a Readable Writing Approach to Text Revision." *Journal of Educational Psychology*, Vol. 74, No. 5, 733-748.
- Fodor, J. A., and M. Garrett. 1967. "Some Syntactic Determinants of Sentential Complexity." *Perception and Psychophysics*, 2: 289-296.
- Fletcher, Charles R. 1981. "Short-Term Memory Processes in Text Comprehension." *Journal of Verbal Learning and Verbal Behavior*, Vol. 20: 564-574.
- Frase, Lawrence, and Donald Fisher. 1977. "The Comprehensibility of Technical Writing." Paper from the NIE Conference on Writing, June, 1977.
- Frase, Lawrence. 1981. "Ethics of Imperfect Measures." *IEEE Transactions on Professional Communication*, No.1, March 1981: 49-50.
- Frazier, Lyn, and Keith Rayner. 1982. "Making and Correcting Errors During Sentence Comprehension: Eye Movements in the Analysis of Structurally Ambiguous Sentences." *Cognitive Psychology*, 14: 178-210.
- Greenberg, J. H. 1966. "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements." In J. H. Greenberg (ed.), *Universals of Language*, 2d ed. Cambridge, Massachusetts, MIT Press.
- Hake, Rosemary L., and Joseph M. Williams. 1981. "Style and its Consequences: Do as I Do, Not as I Say." *College English*, 43, No. 5: 433-451.
- Harris, Albert J., and Milton D. Jacobson. 1979. "A Framework for Readability Research: Moving Beyond Herbert Spencer." *Journal of Reading*, February 1979, 390-398.
- Holmes, V. M. 1979. "Some Hypotheses About Syntactic Processing in Sentence Comprehension." In William E. Cooper and Edward C.T. Walker (eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*. Hillsdale, New Jersey, Lawrence Erlbaum Associates.
- Huckin, T. 1983. "Cognitive Approach to Readability." In Paul V. Anderson, R. John Brockmann, and Carolyn R. Miller (eds.), *New Essays in Technical and Scientific Communication: Research, Theory, Practice*. New York, Baywood Publishing Company, 99-108.
- Isakson, Richard L. 1979. "Cognitive Processing in Sentence Comprehension." *Journal of Educational Research*, Vol. 72, No. 3.
- Jacobson, Milton D. 1965. "Reading Difficulty of Physics and Chemistry Textbooks." *Educational and Psychological Measurement*, 25: 449-457.

- Kintsch, W. 1974. *The Representation of Meaning in Memory*. Hillsdale, New Jersey, Lawrence Erlbaum Associates.
- , 1975. "Assessing Readability." *Reading Research Quarterly*, No. 1, 1974-75: 62-102.
- , 1963. *The Measurement of Readability*. Ames, Iowa, Iowa State University Press.
- , 1974. "Assessing Readability." *Reading Research Quarterly*, 1: 62-102.
- , 1981. "Readability Indexes: Do They Inform or Misinform?" *Information Design Journal*, Vol. 2, No. 3&4: 251-255.
- , 1976. "A Second Look at the Validity of Readability Formulas." *Journal of Reading Behavior*, 8: 129-152.
- Lanham, R. A. 1979. *Revising Prose*. New York, Scribner.
- Levin, Harry, and Eleanor Kaplan. 1971. "Listening, Reading and Grammatical Structure." In David Horton and James Jenkins (eds.), *The Perception of Language*. Columbus, Charles S. Merrill.
- Macdonald, Nina H., Lawrence T. Frase, Patricia S. Gingrich, and Stacey A. Keenan. 1982. "The Writer's Workbench: Computer Aids for Text Analysis." *IEEE Transactions on Communications*, No. 1, January 1982: 105-110.
- Marslen-Wilson, William, and Lorraine Komisarjevsky Tyler. 1980. "The Temporal Structure of Spoken Language Understanding." *Cognition*, 8: 1-71.
- McCall, William, and Lelah Crabbs. 1925. *Standard Test Lessons in Reading*. New York, Bureau of Publications, Teachers College, Columbia University. 1950 Edition, 1961 Edition.
- McNeill, David. 1966. "Speaking of Space." *Science*, 152: 875-880.
- Miller, James R., and Walter Kintsch. 1980. "Readability and Recall of Short Prose Passages: A Theoretical Analysis." *Journal of Experimental Psychology: Human Learning and Memory*, Vol. 6, No. 4: 335-354.
- Pearson, P. D. 1974. "The Effects of Grammatical Complexity on Children's Comprehension, Recall, and Conception of Certain Semantic Relations." *Reading Research Quarterly*, 10: 155-192.
- Pullum, G. K. 1977. "Word Order Universals and Grammatical Relations." In P. Cole and J. M. Sadock (eds.), *Syntax and Semantics (Vol. 8): Grammatical Relations*. New York, Academic Press.

- Rayner, Keith, Marcia Carlson, and Lyn Frazier, 1983. "The Interaction of Syntax and Semantics During Sentence Processing: Eye Movements in the Analysis of Semantically Biased Sentences." *Journal of Verbal Learning and Verbal Behavior*, 22: 358-374.
- Redish, Janice C. 1980. "Readability." In Daniel B. Felker (ed.), *Document Design: A Review of the Relevant Research*. American Institutes for Research.
- Romelhart, D., P. H. Lindsay, and D. Norman. 1972. "A Process Model for Long-Term Memory." In E. Tulving and W. Donaldson (eds.), *Organization of Memory*. New York, Academic Press.
- Selden, Ramsay W. 1977. "Development of a Frequency-Based Measure of Syntactic Difficulty for Estimating Readability." Unpublished doctoral dissertation. University of Virginia.
- Selzer, J. 1983. "What Constitutes a Readable, Technical Style?" In Paul V. Anderson, R. John Brockmann, and Carolyn R. Miller (eds.), *New Essays in Technical and Scientific Communication: Research, Theory, Practice*. New York, Baywood Publishing Company, 71-89.
- Stevens, Kathleen C. 1980. "Readability Formulae and McCall-Crabbs Standard Test Lessons in Reading." *The Reading Teacher*, 413-415.
- Strunk, W. L., and E. B. White. 1959. *The Elements of Style*. New York, MacMillan.
- Templeton Shane, Carolyn Cain, and James O. Miller. 1981. "Reconceptualizing Readability: The Relationship Between Surface and Underlying Structure Analyses in Predicting the Difficulty of Basal Reader Stories." *Journal of Educational Research*, 74: 382-387.
- Williams, Joseph M. 1978. "Nominal and Verbal Style." Unpublished address given at the University of California, Los Angeles, May 25, 1978.
- Wisher, Robert A. 1976. "The Effects of Syntactic Expectations During Reading." *Journal of Educational Psychology*, Vol. 68, No. 5: 597-602.

END

FILMED

5-85

DTIC